基于辨识特征后融合的行人再识别 *

刘　琦[1, 2]，侯　丽[1, 2]

(1. 黄山学院 信息工程学院，安徽 黄山 245041；2. 上海大学 通信与信息工程学院，上海 200444)

摘　要：跨摄像机行人因光照、视角、姿态的差异，会使其外观变化显著，给行人再识别的研究带来严峻挑战。基于多特征融合和距离度量学习技术，提出辨识特征后融合的算法，并将其应用于行人再识别中。首先，对跨摄像机行人样本图像分别提取局部最大出现频次（LOMO）特征和基于显著颜色名称的颜色描述子（SCNCD）特征，表示跨摄像机行人的外观；然后，基于所提取的 LOMO 和 SCNCD 特征，分别去训练跨视图二次判别分析（XQDA）距离度量学习模型，分别获取跨摄像机每对行人每个特征优化的距离；最后，应用最小最大标准化距离融合的算法，获取跨摄像机行人最终的距离，用于跨摄像机行人的匹配。在具有挑战的 VIPeR 和 PRID450S 两个公开数据集上进行实验，实验结果表明所提出的行人再识别算法有效地提高了行人再识别的准确率。

关键词：行人再识别；多特征融合；距离度量学习；距离融合；最小最大标准化

中图分类号：TP391.41　　doi: 10.3969/j.issn.1001-3695.2018.02.0186

Discriminative feature based late fusion for person re-identification

Liu Qi[1, 2], Hou Li[1, 2]

(1. *School of Information Engineering*, *Huangshan University*, *Huangshan* 245041, *China*; 2. *School of Communication & Information Engineering*, *Shanghai University*, *Shanghai* 200444, *China*)

Abstract: Pedestrian may vary greatly in appearance due to differences in illumination, viewpoint, and pose across cameras, which can bring serious challenges in person re-identification. This paper proposed discriminative feature based late fusion using multiple-feature fusion and distance metric learning for person re-identification. Firstly, it expressed human appearance across cameras through local maximal occurrence (LOMO) and salient color names-based color descriptor (SCNCD) extracted from human sample images across cameras. Secondly, it obtained each optimized feature distance of each pair of pedestrians through individual cross-view quadratic discriminant analysis (XQDA) distance metric learning model trained based on LOMO feature and SCNCD feature, respectively. Finally, it obtained the final distance of each pair of pedestrians across cameras through a distance fusion using min-max normalization, which was applied for human match across cameras. Experimental results have shown that the proposed algorithm effectively improves the accuracy of person re-identification on two challenging datasets (VIPeR, PRID450s) .

Key words: person re-identification; multiple-feature fusion; distance metric learning; distance fusion; min-max normalization

0　引言

行人再识别技术，是指让计算机识别不同摄像机下的行人是否是相同的身份，通过行人的外观进行跨摄像机行人的匹配，为摄像机网络自动行人跟踪，以及城市自动视频监控等视频大数据应用提供关键技术支持。然而不同摄像机下的行人会因光照、视角、姿态等方面的差异，导致行人的外观变化显著，如图 1 所示（上、下两行分别为行人再识别基准数据集 VIPeR 中的两个不同摄像机下拍摄的行人图像，同一列为相同的行人），给行人再识别的研究带来严峻挑战。

当前对行人再识别的研究主要集中于两方面[1~3]：一是提取具有辨识力的视觉特征，表示跨摄像机行人的外观；二是寻找判别的距离度量学习算法，优化跨摄像机行人的特征距离。绝大多数行人再识别方法采用多特征融合的方法，改善跨摄像机行人再识别的准确率和鲁棒性[4~7]。Farenzena 等人[4]提出局部特征的对称驱动累积（SDALF）算法来描述跨摄像机行人的外观。SDALF 算法编码了三个互补的行人外观视觉特征，包括通过 HSV 颜色直方图表示的整个彩色内容、通过最大稳定的

收稿日期：2018-02-12；修回日期：2018-03-21　　基金项目：国家自然科学基金资助项目（61704161）；安徽省教育厅自然科学研究项目（KJHS2016B03）；黄山学院横向科研项目（hxkt20170059）；黄山学院校地合作项目（2017XDHZ021）

作者简介：刘琦（1982-），男，辽宁盘锦人，讲师，博士研究生，主要研究方向为计算机视觉、机器学习（liuqi577@126.com）；侯丽（1981-），女，博士，副教授，主要研究方向为视频/图像处理、机器学习.

颜色区域表示的颜色空间排列稳定区域，以及通过周期性的高度结构化的块所表示的具有高熵的局部图形，通过对称和不对称的特性处理视角的变化。Kviatkovsky 等人[5]提出使用颜色不变量（ColorInv）进行行人再识别。ColorInv 在 Log 颜色空间由颜色直方图、协方差描述子、基于部分的形状上下文描述子共同来描述行人外观。基于部分的形状上下文描述子，作为一个不变的形状特征描述子，使用人体目标的不同部分，用于描述可辨识的颜色分布的内在结构分布。Yang 等人[6]提出基于显著颜色名称的颜色描述子（SCNCD）进行行人再识别。将在四个不同的颜色空间(原始的 RGB、归一化的 RGB、l1l2l3 和 HSV）计算的 SCNCD 和颜色直方图进行融合，共同描述跨摄像机行人的外观。Liao 等人[7]提出局部最大出现频次（LOMO）特征表示进行行人再识别。LOMO 由 HSV 颜色直方图和尺度不变局部三元模式纹理描述子组成，分析了人体局部的颜色和纹理特征在同一水平条位置出现的概率，并且最大化这一水平条位置出现的概率，以获得鲁棒的特性表示，有效地处理了跨摄像机行人视角的变化。



图 1　VIPeR 数据集的行人图像

基于上述的特征表示方法，采用标准的距离度量如欧氏距离，进行跨摄像机视域行人匹配时，将会严重影响行人再识别的准确率。为了减轻跨摄像机行人匹配时外观特征的差异，提高行人再识别的准确率，近些年来距离度量学习[7~10]方法广泛应用于行人再识别技术中。受基于似然比检验统计推断的启发，Koestinger 等人[8]提出基于特定标签等值约束的距离度量方法，即 KISSME（保持简单直接的度量）进行行人再识别，该种度量学习方法简单、高效，只需要计算两个小尺寸的不相似对和相似对的协方差矩阵，易于扩展到大规模的数据集中。Pedagadi 等人[9]将无监督的主成分分析（PCA）降维与监督的局部费舍尔辨识分析（LFDA）相结合，提出低维流形距离度量学习框架进行行人再识别。LFDA 在最大化类间分离的同时保存了局部邻域结构，实现了样本数据的多模态分布，并通过广义特征值估计 LFDA 变换。然而，当这个距离度量学习框架应用于相对较小的数据集时，可能产生最具辨识力特征的不良压缩结果，从而影响行人再识别的准确率。为了解决这个问题，通过充分利用核技巧和 LFDA 的优点，Xiong 等人[10]进一步采用核 LFDA(KLFDA)距离度量学习进行行人再识别。 KLFDA 是一种封闭的非线性学习方法，在最大化费舍尔优化准则的同时，使用核技巧处理高维特征向量。该方法在进行数据降维的同时，保留了最具辨识力的特征，并能够充分利用选择核的灵活性来改善行人再识别的准确率。然而该方法在使用非线性核时，其计算速度很慢。Liao 等人[7]提出一种判别度量学习方法，即跨视图二次判别分析（XQDA）应用于行人再识别。XQDA 旨在应用跨摄像机视域训练数据，学习一个可判别的低维子空间，与此同时，在这个低维子空间学习一个优化的距离函数，用于衡量跨摄像机视域行人的特征相似度。

为进一步解决跨摄像机行人外观的显著变化问题，以提高行人再识别的准确率，本文在上述多特征融合和距离度量学习技术的基础上，进一步提出辨识特征后融合的算法，即每个辨识特征先分别经过度量学习，再融合各自的特征距离，应用于行人再识别中。

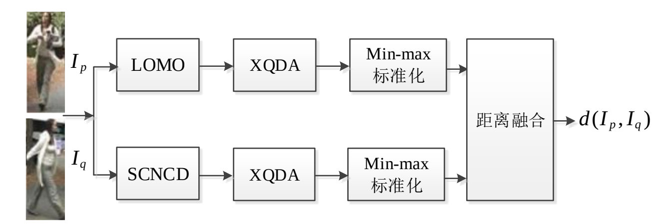


图 2　算法流程

本文提出的行人再识别算法如图 2 所示。首先，从来自不同摄像机下的行人图像中分别提取两个特征，即局部最大出现频次（LOMO）和基于显著颜色名称的颜色描述子（SCNCD）；然后，应用这两个特征分别训练跨视图二次判别分析（XQDA）距离度量学习模型，以获取跨摄像机行人每个特征优化的距离度量；最后，为了平衡每个特征优化的距离度量的贡献，应用最小最大（min-max）标准化算法将每个特征优化的距离度量进行标准化，再应用距离融合的方法，获取跨摄像机行人最终的距离度量，用于跨摄像机行人的匹配，从而进行行人再识别。

1　特征提取

为了有效地处理跨摄像机行人外观的显著变化问题，本文选取对光照和视角具有鲁棒性的视觉特征：LOMO 和 SCNCD，共同来表示跨摄像机行人的外观。

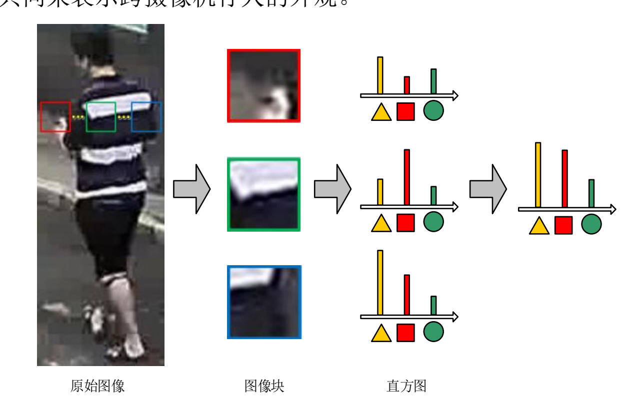


图 3　LOMO 特征提取图解[5]

LOMO 特征提取图解如图 3 所示[7]。采用 10*10 的滑动子窗口表示一幅行人图像的局部区域，在每个子窗口提取 8*8*8-bin 联合 HSV 颜色直方图和 2 个尺度的 SILTP 纹理直方图，然后最大化同一水平位置的所有子窗口每种模式（同一个直方图 bin）的局部发生率，有效地处理了视角变化对行人外观的影响。

此外，采用 3 个尺度的金字塔表示，通过 2*2 局部池化，下采样原始行人图像。上述特征提取步骤重复进行。最后，将每个水平位置所有计算的局部最大的特征向量级联在一起，形成一幅行人图像最终的 LOMO 特征。

考虑颜色特征对跨摄像机光照变化更具辨识力，在上述 LOMO 特征提取的基础上，为了进一步增强光照变化对跨摄像机行人外观的影响，本文在同一幅行人图像中进一步提取 SCNCD 特征。SCNCD 特征提取图解如图 4 所示[6]。共有 M 个索引来描述颜色信息，颜色名称上的数字代表接近某种颜色（这里指红色）的颜色名称集合中的概率分布，只有几个最接近的颜色名称(显著颜色名称)上有非零的概率值。一幅行人图像的 SCNCD 特征就是通过显著颜色名称的概率分布来进行表征。

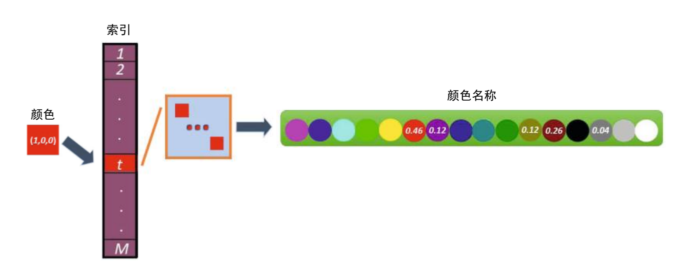


图 4　SCNCD 特征提取图解[4]

2　XQDA 距离度量学习

为了减轻跨摄像机行人外观的显著变化，本文选取快速而有效的 XQDA 距离度量学习方法，获取跨摄像机行人优化的距离度量。

XQDA 应用跨摄像机行人样本数据学习一个低维的特征子空间。同时，在这个低维特征子空间里学习一个优化的距离函数，用于衡量跨摄像机行人的相似度[7]。跨摄像机一对行人样本数据 (x, y) 在低维特征子空间 W 的距离函数 $d_W(x, y)$ 如式(1)所示。

$$d_W(x, y) = (x - y)^{\mathrm{T}} W(\Sigma_{\mathrm{I}}^{'-1} - \Sigma_{\mathrm{E}}^{'-1}) W^{\mathrm{T}} (x - y), \tag{1}$$

$$\begin{cases} \Sigma_I^{'} = W^T \Sigma_I W \\ \Sigma_E^{'} = W^T \Sigma_E W \end{cases} \tag{2}$$

其中：Σ_I 和 Σ_E 分别表示相同行人间的协方差矩阵和不同行人间的协方差矩阵。

基于 LOMO 和 SCNCD 特征，根据跨摄像机行人的训练样本数据集 $\{X,Y\}$，$X = \{x_{mi}\}_{i=1}^{N}$，$Y = \{y_{mi}\}_{i=1}^{N}$（假设每个摄像机有 N 个行人训练样本数据），$m = LOMO, SCNCD$，可分别学习每个特征的核矩阵 $M_m(W) = W_m(\Sigma_{mI}^{'-1} - \Sigma_{mE}^{'-1}) W_m^T$。采用式(1)定义的距离函数可进一步获取每个特征优化的距离度量 $d_{W_m}(x_m, y_m)$。

3　基于 Min-max 标准化距离融合

考虑由 LOMO 和 SCNCD 特征获取的每个特征优化的距离度量范围不同，为了平衡每个特征优化的距离度量的贡献，本文首先采用最小最大（Min-max）标准化算法,将每个特征优化的距离度量先进行标准化到 0～1 后，再进行距离求和运算，作为跨摄像机行人最终的距离度量。

假设基于 LOMO 和 SCNCD 特征优化的距离度量 $d_{W_m}(x_m, y_m)$，$m = LOMO, SCNCD$，经过最小最大（Min-max）标准化后的距离度量用 $d^*_{W_m}(x_m, y_m)$ 来表示。$d^*_{W_m}(x_m, y_m)$ 的计算如式(3)所示。注意，沿着测试集中查询样本的维度来执行最小最大标准化。例如，当一个测试样本和 100 个查询样本时，则通过 100 个距离值来执行最小最大标准化算法。

$$d^*_{W_m}(x_m, y_m) = \frac{d_{W_m}(x_m, y_m) - \min \boldsymbol{d}_{W_m}(x_m, Y)}{\max \boldsymbol{d}_{W_m}(x_m, Y) - \min \boldsymbol{d}_{W_m}(x_m, Y)}, \tag{3}$$

其中：$m = LOMO, SCNCD$；$\boldsymbol{d}_{W_m}(x_m, Y)$ 表示一个距离向量，由一个测试样本和所有查询样本的距离组成；$\min \boldsymbol{d}_{W_m}(x_m, Y)$ 和 $\max \boldsymbol{d}_{W_m}(x_m, Y)$ 分别表示 $\boldsymbol{d}_{W_m}(x_m, Y)$ 中元素的最小值和最大值。明显，$d^*_{W_m}(x_m, y_m)$ 是 $d_{W_m}(x_m, y_m)$ 的线性变换，其值被映射到 0-1 之间。

跨摄像机行人最终的距离度量 $d(I_p, I_q)$ 由上述最小最大（Min-max）标准化后的每个特征的距离度量 $d^*_{W_m}(x_m, y_m)$ 求和来计算，如式(4)所示。

$$d(I_p, I_q) = \sum_m d^*_{W_m}(x_m, y_m). \tag{4}$$

4　实验结果

采用 Windows 7 64 位操作系统 Intel i7 处理器的联想电脑，应用 MATLAB2017a 软件，在具有挑战性的 2 个公开数据集：VIPeR[11]、PRID450S[12]进行实验测试，估计本文所提出的行人再识别算法的累计匹配特性（CMC）。VIPeR 和 PRID450S 两个数据集的简要介绍如表 1 所示。本文随机选取行人数的一半作为训练样本集，另一半作为测试样本集。训练集用于学习不同特征的核矩阵 $M_m(W_m)$。而测试集用于衡量跨摄像机一对行人样本的相似度。为了使实验结果稳定、可靠，当计算 CMC 曲线时，每个实验重复 10 次取平均识别率。

表 1　VIPeR 和 PRID450S 数据集简要介绍.

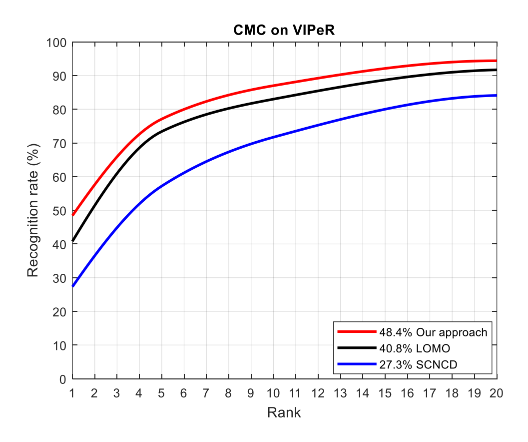
数据集	VIPeR	PRID450S
图像数	1264	900
行人数	632	450

训练集中图像数	316	225
摄像机视图数	2	2
每个视角每个行人的图像数	1	1

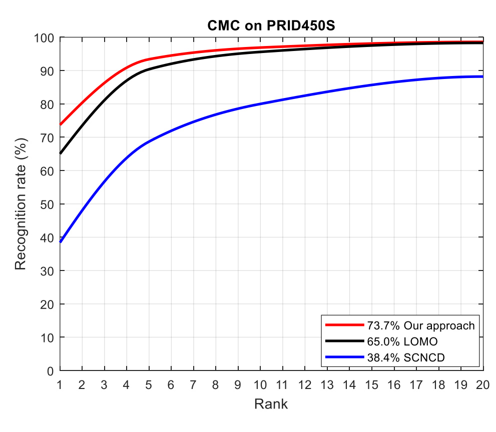
在实验中，将从特征敏感度分析和与其他先进的算法比较两方面展示实验结果：

4.1 特征敏感度分析

采用相同的 XQDA 距离度量学习方法。为了分析 LOMO、SCNCD 特征贡献敏感度，本文以 LOMO、SCNCD 以及本文所提出的方法（图 2）分别在 VIPeR 和 PRID450S 两个数据集下进行实验测试。图 5(a)(b)分别给出了 LOMO、SCNCD，以及本文提出的方法分别在 VIPeR 和 PRID450S 两个数据集下的 CMC 性能比较结果。



(a) VIPeR



(b) PRID450S

图 5 不同特征在 VIPeR、PRID450S 数据集的 CMC 性能比较

由图 5 可见，本文提出的方法最具辨识力，在 VIPeR、PRID450S 两个数据集中排序为 1 时的识别率分别为 48.4%、73.7%。LOMO 特征比 SCNCD 特征更具辨识力，LOMO 特征在 VIPeR、PRID450S 两个数据集中排序为 1 时的识别率分别为 40.8%、65.0%，而 SCNCD 特征在 VIPeR、PRID450S 两个数据集中排序为 1 时的识别率分别为 27.3%、38.4%。显然，本文提出的基于辨识特征后融合的行人再识别方法能够有效地提高行人再识别的准确率。

4.2 与其他先进的算法比较

本文分别在 VIPeR 和 PRID450S 两个数据集下，与近几年提出的先进的行人再识别算法进行 CMC 性能比较。表 2、3 分别给出了相应的 CMC 性能比较结果。由表 2、3 可知，本文提出的方法（图 2）在 VIPeR 和 PRID450S 两个数据集下，明显高于所比较的近几年提出的先进的行人再识别算法，可见辨识特征的后融合算法有助于提高行人再识别的性能。

表 2 VIPeR 数据集近几年公开发表的行人再识别算法的识别率(只列出了排序为 1、5、10、20 的累积匹配分数) /%

VIPeR 排序	1	5	10	20
Our approach	**48.4**	**77.1**	**87.0**	**94.4**
MR [13]	43.0	75.2	87.3	94.8
LSSCDL [14]	42.7	–	84.3	91.9
TSR [15]	41.6	71.9	86.2	95.1
MED_VL[16]	41.1	71.7	83.2	91.7
DRML[17]	41.1	–	79.1	91.3
LOMO[7]	40.0	–	80.5	91.1
ECM [18]	38.9	67.8	78.4	88.9
SCNCD[19]	37.8	68.5	81.2	90.4
CMWCE [20]	37.6	68.1	–	90.2
CSL [21]	34.8	68.7	82.3	91.8
KPLS [22]	33.1	67.7	81.0	90.8

表 3 PRID450S 数据集近几年公开发表的行人再识别算法的识别率(只列出了排序为 1、5、10、20 的累积匹配分数) /%

PRID450s 排序	1	5	10	20
Our approach	**73.7**	**93.4**	**96.9**	**98.6**
LSSCDL[14]	60.5	–	88.6	93.6
DRML[17]	56.4	–	82.2	90.2
MR [13]	55.4	79.3	87.8	93.9
KPLS [22]	52.8	82.1	90.0	95.4
MED_VL[16]	45.9	73.0	82.9	91.1
TSR [15]	44.9	71.7	77.5	86.7
CSL [21]	44.4	71.6	82.2	89.8
ECM [18]	41.9	66.3	76.9	84.9
SCNCD [19]	41.6	68.9	79.4	87.8
CMWCE [20]	40.6	–	80.7	90.0

5 结束语

本文基于 LOMO、SCNCD 特征和 XQDA 距离度量学习的基础上，提出辨识特征的后融合方法，并应用于行人再识别中。在具有挑战的 VIPeR 和 PRID450S 两个行人再识别基准数据集上进行实验，实验结果展示了本文提出的行人再识别算法的先进性和有效性。

参考文献：

[1] Wang Xiaogang. Intelligent multi-camera video surveillance: a review [J]. Pattern Recognition Letters, 2013, 34 (1): 3-19

[2] Yu Tianshu, Wang Ruisheng. Enhancing scene parsing by transferring structures via efficient low-rank graph matching [C]// Proc of ACM

SIGSPATIAL International Conference on Geographic Information Systems. 2016: 1-9.

[3] Zhang Xin, Ding Meng, Fan Guoliang. Video-based human walking estimation using joint gait and pose manifolds [J]. IEEE Trans on Circuits and Systems for Video Technology, 2017, 27 (7): 1540-1554.

[4] Farenzena M, Bazzani L, Perina A, *et al.* Person re-identification by symmetry-driven accumulation of local features [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 2360-2367.

[5] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person re-identification [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (7): 1622-1634.

[6] Yang Yang, Yang Jimei, Yan Junjie, *et al.* Salient color names for person re-identification [C]// Proc of European Conference on Computer Vision. 2014: 536-551.

[7] Liao Shengcai, Hu Yang, Zhu Xiangyu, *et al.* Person re-identification by local maximal occurrence representation and metric learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2197-2206.

[8] Koestinger M, Hirzer M, Wohlhart P, *et al.* Large scale metric learning from equivalence constraints [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2288-2295.

[9] Pedagadi S, Orwell J, Velastin S, *et al.* Local fisher discriminant analysis for pedestrian re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3318-3325.

[10] Xiong Fei, Gou Mengran, Camps O, *et al.* Person re-identification using kernel-based metric learning methods [C]// Proc of European Conference on Computer Vision. 2014: 1-16.

[11] Gray D, Brennan S, Tao Hai. Evaluating appearance models for recognition, reacquisition, and tracking [C]// Proc of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance. 2007: 1-7.

[12] Roth P M, Hirzer M, Köstinger M, *et al.* Mahalanobis distance learning for person re-identification [M]// Person Re-Identification. London: Springer, 2014: 247-267.

[13] Chen Yingcong, Zheng Weishi, Lai Jianhuang. Mirror representation for modeling view-specific transform in person re-identification [C]// Proc of International Joint Conference on Artificial Intelligence. 2015: 3402-3408.

[14] Zhang Ying, Li Baohua, Lu Huchuan, *et al.* Sample-specific SVM learning for person re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1278-1287.

[15] Shi Zhiyuan, Hospedales T M, Xiang Tao. Transferring a semantic representation for person re-identification and search [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4184-4193.

[16] Yang Yang, Lei Zhen, Zhang Shifeng, *et al.* Metric embedded discriminative vocabulary learning for high-level person representation [C]// Proc of AAAI Conference on Artificial Intelligence. 2016: 3648-3654.

[17] Yao Wenbin, Weng Zhenyu, Zhu Yuesheng. Diversity regularized metric learning for person re-identification [C]// Proc of IEEE International Conference on Image Processing. 2016: 4264-4268.

[18] Liu Xiaokai; Wang Hongyu; Wu Yi, *et al.* An ensemble color model for human re-identification [C]// Proc of IEEE Winter Conference on Applications of Computer Vision. 2015: 868-875.

[19] Yang Yang, Yang Jimei, Yan Junjie, *et al.* Salient color names for person re-identification [C]// Proc of European Conference on Computer Vision. 2014: 536-551.

[20] Yang Yang, Liao Shengcai, Lei Zhen, *et al.* Color models and weighted covariance estimation for person re-identification [C]// Proc of International Conference on Pattern Recognition. 2014: 1874-1879.

[21] Shen Yang, Lin Weiyao, Yan Junchi, *et al.* Person re-identification with correspondence structure learning [C]// Proc of IEEE International Conference on Computer Vision. 2015: 3200-3208.

[22] Prates R, Oliveira M, Schwartz W R. Kernel partial least squares for person re-identification [C]// Proc of IEEE International Conference on Advanced Video and Signal Based Surveillance. 2016: 249-255.